AD-A104 938 TEXAS A AND M UNIV COLLEGE STATION INST OF STATISTICS F/G 12/1
ON SOME NUMERICAL PROPERTIES OF ARMA PARAMETER ESTIMATION PROCE--ETC(U)
JUL 81 H J NEWTON DAAG29-80-C-0070
UNCLASSIFIED TR-N-25 NL

TEXAS A&M UNIVERSITY
COLLEGE STATION, TEXAS 77843

INSTITUTE OF STATISTICS
Phone 713 - 845-3141

ON SOME NUMERICAL PROPERTIES OF ARMA PARAMETER ESTIMATION PROCEDURES

by H. Joseph Newton

Institute of Statistics

Texas A&M University

Technical Report No. N-25

July 1981

Texas A & M Research Foundation
Project No. 4226T

"Robust Statistical Data Analysis and Modeling"

Sponsored by the Office of Naval Research

Professor Emanuel Parzen, Principal Investigator

Approved for public release; distribution unlimited.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: TR-N-25	2. GOVT ACCESSION NO.: AD A104938
3. RECIPIENT'S CATALOG NUMBER:	
4. TITLE (and Subtitle): On Some Numerical Properties of ARMA Parameter Estimation Procedures,	5. TYPE OF REPORT & PERIOD COVERED: Technical
	6. PERFORMING ORG. REPORT NUMBER:
7. AUTHOR(s): H. Joseph Newton	8. CONTRACT OR GRANT NUMBER(s): DAAG29-80-C-0070 ~~ONR N0001481MP10001~~
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Texas A&M University, Institute of Statistics, College Station, TX 77843	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS:
11. CONTROLLING OFFICE NAME AND ADDRESS: U. S. Army Research Office, P. O. Box 12277, Research Triangle Park, NC 27709	12. REPORT DATE: 11 Jul 81
	13. NUMBER OF PAGES: 6 (six)
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office):	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

NA

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

ARMA models, Maximum likelihood estimation, Minimum mean square error prediction, Kalman filter algorithm, Modified Cholesky decomposition algorithm.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This paper reviews the algorithms used by statisticians for obtaining efficient estimators of the parameters of a univariate autoregressive moving average (ARMA) time series. The connection of the estimation problem with the problem of prediction is investigated with particular emphasis on the Kalman filter and modified Cholesky decomposition algorithms. A result from prediction theory is given which provides a significant reduction in the computations needed in Ansley's (1979) estimation procedure. Finally it is pointed out that there are many useful facts in the literature of control theory that need to be investigated by statisticians interested in estimation and prediction problems in linear time series models.

DD FORM 1 JAN 79 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ON SOME NUMERICAL PROPERTIES OF ARMA PARAMETER ESTIMATION PROCEDURES

H. Joseph Newton
Institute of Statistics,
Texas A&M University

Abstract

This paper reviews the algorithms used by statisticians for obtaining efficient estimators of the parameters of a univariate autoregressive moving average (ARMA) time series. The connection of the estimation problem with the problem of prediction is investigated with particular emphasis on the Kalman filter and modified Cholesky decomposition algorithms. A result from prediction theory is given which provides a significant reduction in the computations needed in Ansley's (1979) estimation procedure. Finally it is pointed out that there are many useful facts in the literature of control theory that need to be investigated by statisticians interested in estimation and prediction problems in linear time series models.

Key Words

ARMA models, Maximum likelihood estimation, Minimum mean square error prediction, Kalman filter algorithm, Modified Cholesky decomposition algorithm

1. INTRODUCTION

Let $\{\varepsilon(t), t\varepsilon Z\}$, Z the set of integers, be a white noise time series of uncorrelated zero mean random variables having common variance σ^2. Then the time series $\{Y(t), t\varepsilon Z\}$ satisfying

$$Y(t) = -\sum_{j=1}^{p} \alpha(j)Y(t-j)+\varepsilon(t)+\sum_{k=1}^{q} \beta(k)\varepsilon(t-k) \quad (1.1)$$

is called an autoregressive moving average process of order p and q (ARMA(p,q)). If p=0 then $Y(\cdot)$ is a moving average process of order q, MA(q), while if q=0 $Y(\cdot)$ is an autoregressive process of order p, AR(p).

Defining $\alpha(0)=\beta(0)=1$ and the complex valued polynomials

$$g(z)=\sum_{j=0}^{p} \alpha(j)z^j \ , \ h(z)=\sum_{k=0}^{q} \beta(k)z^k \ ,$$

we can write (1.1) as

$$\sum_{j=0}^{p} \alpha(j)Y(t-j)=\sum_{k=0}^{q} \beta(k)\varepsilon(t-k)$$

or

$$g(L)Y(t)=h(L)\varepsilon(t),$$

where L is the lag or back shift operator, $L^jY(t)=Y(t-j)$, $j\varepsilon Z$. If the zeros of $g(\cdot)$ are outside the unit circle then $Y(\cdot)$ is stationary, i.e. it can be written as an infinite order moving average process, while if the zeros of $h(\cdot)$ are outside the unit circle then $Y(\cdot)$ is invertible, i.e. it can be written as an infinite order autoregressive process.

The ARMA model has been very useful in analyzing time series data. Given a sample realization $\underset{\sim}{Y}_T^T=(Y(1),\ldots,Y(T))$ from $Y(\cdot)$ one seeks estimators $\hat{\underset{\sim}{\alpha}}=(\hat{\alpha}(1),\ldots,\hat{\alpha}(p))^T$, $\hat{\underset{\sim}{\beta}}=(\hat{\beta}(1),\ldots,\hat{\beta}(q))^T$, and $\hat{\sigma}^2$ of the parameters $\underset{\sim}{\alpha}$, $\underset{\sim}{\beta}$, and σ^2, as well as memory-t, horizon-h, minimum mean square error linear predictors and prediction variances $Y(t+h|t)$, $\sigma^2_{t,h}=E\{Y(t+h)-Y(t+h|t)\}^2$ of $Y(t+h)$ given $Y(1),\ldots,Y(t)$ for a variety of memories and horizons $t=t_1,\ldots,t_2$, $h=h_1,\ldots,h_2$. Thus $Y(t+h|t)=\underset{\sim}{\lambda}_{t,h}^T\underset{\sim}{Y}_t$ where $\underset{\sim}{\lambda}_{t,h}$ minimizes

$$S(\underset{\sim}{\ell})=E\{Y(t+h)-\underset{\sim}{\ell}^T\underset{\sim}{Y}_t\}^2 \ ,$$

and $\sigma^2_{t,h}=S(\underset{\sim}{\lambda}_{t,h})$. This gives (see Whittle (1963), p.47) that $\underset{\sim}{\lambda}_{t,h}$ satisfies the normal equations

$$\Gamma_{Y,t}\underset{\sim}{\lambda}_{t,h}=\underset{\sim}{r}_{t,h} \ ,$$

while

$$\sigma^2_{t,h}=R_Y(0)-\underset{\sim}{r}^T_{t,h}\,\Gamma^{-1}_{Y,t}\underset{\sim}{r}_{t,h},$$

where $\underset{\sim}{r}^T_{t,h}=(R_Y(t+h-1),\ldots,R_Y(h))$ with $R_Y(v)=E(Y(t)Y(t+v))$, $v\varepsilon Z$, and $\Gamma_{Y,t}$ is the $(t\times t)$ Toeplitz covariance matrix of $\underset{\sim}{Y}_t$, i.e. $\Gamma_{Y,t}$ has (j,k)th element $R_Y(k-j)$. We can thus define $\Gamma_{Y,t}$ by its first row; $\Gamma_{Y,t}=\text{TOEPL}(R_Y(0),\ldots,R_Y(t-1))$.

The usual assumption made for doing estimation and prediction in ARMA models is that $\varepsilon(\cdot)$ (and thus $Y(\cdot)$) is a Gaussian process. Thus the maxi-

mum likelihood estimators $\hat{\underline{\alpha}}, \hat{\underline{\beta}}$, and $\hat{\sigma}^2$ maximize the Gaussian likelihood

$$L(\underline{\alpha}, \underline{\beta}, \sigma^2 / \underline{Y}_T) = (2\pi)^{-T/2} |\Gamma_{Y,T}|^{-\frac{1}{2}} \exp(-\tfrac{1}{2} \underline{Y}_T^T \Gamma_{Y,T}^{-1} \underline{Y}_T),$$

while $Y(t+h|t)$ and $\sigma^2_{t,h}$ are the conditional mean and variance of $Y(t+h)$ given $Y(1),\ldots,Y(t)$.

In this paper we 1) review attempts by statisticians to obtain $\hat{\underline{\alpha}}, \hat{\underline{\beta}}$, and $\hat{\sigma}^2$ (or estimators asymptotically equivalent to them), 2) show how recent methods are closely connected with finding predictors, and 3) propose an improvement of Ansley's (1979) estimation procedure. This procedure is currently regarded, at least in many situations, as the most numerically efficient available. Finally we will be able to compare the two most popular methods available.

2. APPROXIMATE METHODS

There have been three basic types of procedures used to estimate the parameters of the ARMA model: 1) estimators that are derived heuristically and then shown to be asymptotically equivalent to the MLE, 2) estimators obtained by maximizing a function asymptotically equivalent to the likelihood L, and 3) directly maximizing L. The procedures have evolved as both computing software and hardware have improved.

As an example of a type 1 procedure we consider Hannan's (1969) method. The method consists of 1) finding consistent initial estimators $\hat{\underline{\alpha}}^{(0)}$ and $\hat{\underline{\beta}}^{(0)}$, 2) performing an alternating procedure of the form $\hat{\underline{\alpha}}^{(0)} \to \hat{\underline{\beta}}^{(1)}, \hat{\underline{\beta}}^{(1)} \to \hat{\underline{\alpha}}^{(1)}, \hat{\underline{\alpha}}^{(1)} \to \hat{\underline{\beta}}^{(2)}$, 3) combining $\hat{\underline{\beta}}^{(1)}, \hat{\underline{\beta}}^{(2)}$ to obtain asymptotically efficient $\tilde{\underline{\beta}}$, 4) $\tilde{\underline{\beta}} \to \tilde{\underline{\alpha}}$, and 5) possibly iterating the process by returning to (2) with $\tilde{\underline{\alpha}}$ replacing $\hat{\underline{\alpha}}^{(0)}$.

The initial estimators $\hat{\underline{\alpha}}^{(0)}$ and $\hat{\underline{\beta}}^{(0)}$ are obtained by noting the following facts about the ARMA model:

1) $$\sum_{j=0}^{p} \alpha(j) R_Y(j-v) = 0 \ , \ v = q+1, \ldots, q+p \qquad (2.1)$$

2) By writing

$$g(L)Y(t) = h(L)\varepsilon(t) = X(t), \qquad (2.2)$$

we note that $X(\cdot) \sim MA(q, \underline{\beta}, \sigma^2)$, i.e. $X(\cdot)$ is an MA(q) process with parameters $\underline{\beta}$ and σ^2. Thus by the first equality in (2.2) we have

$$R_X(v) = \sum_{j,k=0}^{p} \alpha(j)\alpha(k) R_Y(j+v-k), v \in Z, \qquad (2.3)$$

while from the second equality we have

$$R_X(v) = \begin{cases} \sigma^2 \sum_{k=0}^{q-|v|} \beta(k)\beta(k+|v|) \ , & |v| \le q \\ 0 \ , & |v| > q \ . \end{cases} \qquad (2.4)$$

Then defining the consistent sample autocovariances

$$R_T(v) = \frac{1}{T} \sum_{t=1}^{T-|v|} Y(t) Y(t+|v|) \ , \quad |v| < T \ ,$$

$\hat{\underline{\alpha}}^{(0)}$ is found by solving (2.1) with $R_T(v)$ replacing $R_Y(v)$, while $\hat{\underline{\beta}}^{(0)}$ is obtained by first using $\hat{\underline{\alpha}}^{(0)}$ and $R_T(\cdot)$ in (2.3) to obtain consistent estimators $R_X^{(0)}(0), \ldots, R_X^{(0)}(q)$ which are then used in (2.4) to get $\hat{\underline{\beta}}^{(0)}$ via Wilson's (1969) or Bauer's (1955) algorithm.

Steps 2 and 4 of Hannan's method are performed in the frequency domain by noting from (2.2) that the spectral density functions $f_Y(\cdot)$, $f_X(\cdot)$, and $f_\varepsilon(\cdot)$ of $Y(\cdot)$, $X(\cdot)$, and $\varepsilon(\cdot)$ are related by (since $f_\varepsilon(w) = \sigma^2/2\pi$, $w \in [0, 2\pi]$)

$$|g(e^{iw})|^2 f_Y(w) = \frac{\sigma^2}{2\pi} |h(e^{iw})|^2 = f_X(w), \qquad (2.5)$$

Then we can rewrite (2.5) as, dropping arguments for convenience,

$$\frac{|g|^2 f_Y}{f_X^2} = \frac{4\pi^2/\sigma^2}{2\pi} \frac{1}{|h|^2} \qquad (2.6)$$

$$\frac{f_Y}{|h|^2} = \frac{\sigma^2}{2\pi} \frac{1}{|g|^2} \qquad (2.7)$$

$$\frac{(2\pi/\sigma^2)^2 |g|^2 f_Y}{|h|^4} = \frac{4\pi^2/\sigma^2}{2\pi} \frac{1}{|h|^2} \ . \qquad (2.8)$$

Thus the right hand sides of (2.6), (2.7), and (2.8) are in the form of an autoregressive spectral density, and thus for $v \ge 0$,

$$\sum_{j=0}^{q} \beta(j) \int_0^{2\pi} \frac{|g|^2 f_Y}{f_X^2} e^{i(j-v)w} dw = \delta_v \frac{4\pi^2}{\sigma^2} \ , \qquad (2.9)$$

$$\sum_{j=0}^{p} \alpha(j) \int_0^{2\pi} \frac{f_v}{|h|^2} e^{i(j-v)w} dw = \delta_v \sigma^2 \ , \qquad (2.10)$$

$$\sum_{j=0}^{q} \beta(j) \int_0^{2\pi} \frac{|g|^2 f_Y}{|h|^4} e^{i(j-v)w} dw = \delta_v \sigma^2 \ . \qquad (2.11)$$

Now f_Y can be estimated by the periodogram

$$f_T(w) = \frac{1}{2\pi} \sum_{|v|<T} R_T(v) e^{-ivw} \ ,$$

and since $X(\cdot) \sim MA(q, \underline{\beta}, \sigma^2)$ we have that

$$f_X(w) = \frac{1}{2\pi} \sum_{|v| \le q} R_X(v) e^{-ivw} \ .$$

Then for example, in an obvious notation, the step $\hat{\alpha}^{(0)} \to \hat{\beta}^{(1)}$ is done by solving (2.9) with $g^{(0)}$, f_T, and $f_X^{(0)}$ replacing g, f_Y, and f_X and the fast Fourier transform used to calculate a rectangular sum approximation to the integrals needed. Then (2.10) is used for $\hat{\beta}^{(1)} \to \hat{\alpha}^{(1)}$ and $\hat{\beta} \to \hat{\alpha}$ while (2.11) is used for $\hat{\alpha}^{(1)} \to \hat{\beta}^{(2)}$. Finally the combination of $\hat{\beta}^{(1)}$ and $\hat{\beta}^{(2)}$ is done using an analogy with two-stage least squares.

In a pivotal article, Akaike (1973) noted that Hannan's method was the same as one step in directly maximizing L using the Newton-Raphson procedure with an approximation to the Hessian matrix of L. This observation led many researchers to turn their attention to other possible approaches to directly maximize L. We will consider two of these in the next section.

As an example of a type 2 estimation procedure we consider the method of Box and Jenkins (1970). In large samples the term $\exp(-\frac{1}{2}\underset{\sim}{Y}_T^T\Gamma_{Y,T}^{-1}\underset{\sim}{Y}_T)$ dominates the term $|\Gamma_{Y,T}|^{-\frac{1}{2}}$. Thus Box and Jenkins suggest maximizing

$$L'(\underset{\sim}{\alpha}, \underset{\sim}{\beta}, \sigma^2|\underset{\sim}{Y}_T) = \exp(-\underset{\sim}{Y}_T^T\Gamma_{Y,T}^{-1}\underset{\sim}{Y}_T)$$

which can be done in a nonlinear regression framework. Thus it can be shown that

$$\underset{\sim}{Y}_T^T\Gamma_{Y,T}^{-1}\underset{\sim}{Y}_T = \sum_{t=-\infty}^{T} [\varepsilon(t)]^2 \qquad (2.12)$$

where $[\varepsilon(t)] = E(\varepsilon(t)|\underset{\sim}{Y}_T, \underset{\sim}{\alpha}, \underset{\sim}{\beta})$. Thus the Box-Jenkins procedure consists of replacing $-\infty$ by $-Q$ in (2.12) and approximating $[\varepsilon(t)]$ by backforecasting.

It is generally agreed that in the case of large T and/or zeros of h(z) far from the unit circle that estimation methods of type 1 and type 2 give very reasonable results. However if one is faced with a situation where the above conditions are not satisfied, then the numerical properties of the procedures nullify the benefit of their simplicity. Thus iterations often fail to converge in a reasonable time and systems of equations that must be solved become highly ill-conditioned. Thus there has been a need for more numerically stable, exact maximum likelihood methods.

3. EXACT MAXIMUM LIKELIHOOD PROCEDURES

With the advent of iterative nonlinear optimization procedures which require only a starting value and evaluation of the function to be optimized for given values of its arguments (see Dennis and More (1977), for example), the most recent procedures for ARMA parameter estimation have centered on evaluating L for given values of $\underset{\sim}{\alpha}$, $\underset{\sim}{\beta}$, and σ^2.

In this section we consider two such methods for evaluating L: 1) using the Kalman filter algorithm and 2) using covariance matrix decomposition methods, particularly the Cholesky decomposition algorithm. The basic purpose of these algorithms is to obtain the one step ahead prediction errors $e(t) = Y(t) - Y(t|t-1)$ and prediction variances $\sigma_t^2 = \sigma_{t-1,1}^2$, $t = 1,\ldots, T$ since one can easily show that

$$|\Gamma_{Y,T}| = \prod_{t=1}^{T} \sigma_t^2, \quad \underset{\sim}{Y}_T^T\Gamma_{Y,T}^{-1}\underset{\sim}{Y}_T = \sum_{t=1}^{T} \frac{e^2(t)}{\sigma_t^2}.$$

To summarize, the two methods currently regarded as most numerically efficient for maximizing L consist of two stages:

1) Find initial estimators as described above for Hannan's method.

2) Use an iterative, derivative free nonlinear optimization algorithm to find the maximum likelihood estimators, using either the Kalman filter algorithm (Akaike (1974), Harvey and Phillips (1979), Jones (1980), Pearlman (1980), or Gardner, Harvey, and Phillips (1980), for example) or the Cholesky decomposition algorithm (see Pagano and Parzen (1973), Pagano (1976), Phadke and Kedem (1978), Ansley (1979), Newton (1980), Newton and Pagano (1981), for example) to find the $e(t)$ and σ_t^2 needed to evaluate L.

Kalman Filter Algorithm

Consider the following two equations:

$$\underset{\sim}{Y}_{t+1} = H_t\underset{\sim}{Y}_t + G_t\underset{\sim}{U}_t \quad \text{(State Equation)}$$

$$\underset{\sim}{X}_{t+1} = S_{t+1}\underset{\sim}{Y}_{t+1} + \underset{\sim}{V}_{t+1} \quad \text{(observation equation)}$$

where the $\underset{\sim}{Y}$'s are unobservable N-vectors, the $\underset{\sim}{X}$'s are observable M-vectors, H_t, G_t, and S_t are known (NxN), (NxL), and (MxN) matrices, and $\underset{\sim}{U}_t$, $\underset{\sim}{V}_t$ are independent zero mean L and N dimensional white noise series with known covariance matrices Q_t and R_t.

Then given initial values $\underset{\sim}{Y}_0$ and $P_0 = \text{var}(\underset{\sim}{Y}_0)$ one finds the $\underset{\sim}{Y}_t$'s and $P_t = \text{cov}(\underset{\sim}{Y}_t)$ by the Kalman filter algorithm

$$\underset{\sim}{Y}_{t+1} = \underset{\sim}{Z}_{t+1} - K_{t+1}(S_{t+1}\underset{\sim}{Z}_{t+1} - \underset{\sim}{X}_{t+1})$$

$$P_{t+1} = W_{t+1} - K_{t+1}S_{t+1}W_{t+1}$$

$$\underset{\sim}{Z}_{t+1} = H_t\underset{\sim}{Y}_t, \quad K_{t+1} = W_{t+1}S_{t+1}^T \times (S_{t+1}W_{t+1}S_{t+1}^T + R_{t+1})^{-1}$$

$$W_{t+1} = H_tP_tH_t^T + G_tQ_tG_t^T. \qquad (3.1)$$

A number of authors in the statistical litera-

ture have pointed out that a simple application of this algorithm to ARMA models can be made to obtain the $e(t)$ and σ_t^2. Thus since

$$Y(t+j)|t+1) = \begin{cases} Y(t+j|t) + g_j \varepsilon(t+1), \\ \qquad j = 1, \ldots, m-1 \\ \sum_{k=1}^{p} \alpha(k) Y(t+m-k|t) + g_m \varepsilon(t+1), \\ \qquad j = m \end{cases}$$

where $g_1 = 1$ and $g_j = \beta(j-1) + \sum_{k=1}^{j-1} \alpha(k) g_{j-k}, j \geq 2$

and $m = \max(p, q+1)$, we can write the equation of state

$$\underset{\sim}{Y}_{t+1} = H \underset{\sim}{Y}_t + \underset{\sim}{g}\, \varepsilon(t+1) ,$$

where $\underset{\sim}{g}^T = (g_1, g_2, \ldots, g_m)$, and

$$H = \begin{bmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ . & . & . & & . \\ . & . & . & & . \\ . & . & . & & . \\ 0 & 0 & 0 & \ldots & 1 \\ \alpha(m) & . & . & \ldots & \alpha(1) \end{bmatrix} ,$$

where $\alpha(j) = 0$ if $j > p$.

Further, since $Y(t+1|t+1) = Y(t+1)$, we can write the observational equation

$$\underset{\sim}{X}_{t+1} = \underset{\sim}{S}^T \underset{\sim}{Y}_{t+1} + \underset{\sim}{V}_{t+1} ,$$

where $\underset{\sim}{X}_{t+1} = Y(t+1)$, $\underset{\sim}{S}^T = (1, 0, \ldots, 0)$, and $\underset{\sim}{V}_{t+1}$ is an observational error random variable.

Cholesky Decomposition Algorithm

An $(n \times n)$ symmetric matrix A_n is positive definite if and only if it can be written

$$A_n = L_{A,n} D_{A,n} L_{A,n}^T , \qquad (3.2)$$

where $L_{A,n}$ is a unit lower triangular matrix and $D_{A,n}$ is a diagonal matrix with positive diagonal elements. This is the modified Cholesky decomposition of A_n. We note that we can also write the Cholesky decomposition $A_n = M_{A,n} M_{A,n}^T$ where $M_{A,n} = L_{A,n} D_{A,n}^{\frac{1}{2}}$ but we will consider exclusively the modified decomposition.

The decomposition (3.2) is unique and nested, i.e. $L_{A,j}$, $D_{A,j}$, $L_{A,j}^{-1}$, $D_{A,j}^{-1}$ are the $(j \times j)$ principal minors of $L_{A,n}$, $D_{A,n}$, $L_{A,n}$, and $D_{A,n}^{-1}$ respectively for $j \leq n$. Thus we can denote the (j,k)th element of $L_{Y,t}$ and $D_{Y,t}$ in the decomposition of the $(K \times K)$ covariance matrix $\Gamma_{Y,K}$ of a time series $Y(\cdot)$ as $L_{Y,j,k}$ and $D_{Y,j,k}$ respectively for $t \leq K$. If $Y(\cdot)$ is a purely nondeterministic covariance stationary time series with spectral density function $f(\cdot)$, i.e.

$$\sigma_\infty^2 = 2\pi \exp\{\frac{1}{2\pi} \int_0^{2\pi} \log f(w) dw\} > 0 ,$$

then $e(1) = Y(1)$, while

$$e(t) = Y(t) - \sum_{k=1}^{t-1} L_{Y,t,t-k} e(t-k), \quad t \geq 2$$

$$\sigma_t^2 = D_{Y,t,t} ,$$

and in fact (Newton and Pagano (1981))

$$Y(t+h|t) = \sum_{k=h}^{t+h-1} L_{Y,t+h,t+h-k} e(t+h-k) \qquad (3.3)$$

$$\sigma_{t,h}^2 = \sum_{k=0}^{h-1} L_{Y,t+h,t+h-k}^2 D_{Y,t+h-k,t+h-k} . \qquad (3.4)$$

Further,

$$\lim_{K \to \infty} L_{Y,K,K-j} = \beta_\infty(j), \; j \geq 0 \qquad (3.5)$$

$$\lim_{K \to \infty} D_{Y,K,K} = \sigma_\infty^2 , \qquad (3.6)$$

where the $\beta_\infty(\cdot)$ are the coefficients in the $MA(\infty)$ representation of $Y(\cdot)$. While these facts appear to be known in the literature of control theory they do not appear to exist in the statistical literature, at least in the general form of equations (3.3) through (3.6).

Now major simplifications of these equations occur when $Y(\cdot)$ is an ARMA process. Thus (Pagano and Parzen (1973), Pagano (1976), Newton (1980)) if $Y(\cdot)$ is an $MA(q)$ we have that $\Gamma_{Y,j,k} = 0$ if $|j-k| > 0$ and thus $L_{Y,j,k} = 0$ if $j - k > 0$. Further, Pagano and Parzen (1973) state in an attempt to find $Y(t+h|t)$ and $\sigma_{t,h}^2$ that if $Y(\cdot) \sim ARMA(p, q, \underset{\sim}{\alpha}, \underset{\sim}{\beta}, \sigma^2)$, and one forms the series $X(t) = \sum_{j=0}^{p} \alpha(j) Y(t-j)$, $t > p$, then $X(p+1), \ldots, X(T)$ is a realization from an $MA(q, \underset{\sim}{\beta}, \sigma^2)$ process and one can combine the $MA(q)$ prediction algorithm with an $AR(p)$ prediction algorithm. Ansley (1979) makes this same transformation of $Y(\cdot)$ to $X(\cdot)$ to get the $e(t)$ and σ_t^2 necessary to evaluate L.

The methods of Pagano and Parzen (1973) and Ansley (1979) can be summarized theoretically by combining the equations (3.3) through (3.6) above with the following theorem due to Newton and Pagano (1981).

Theorem

Let $Y(\cdot) \sim ARMA(p, q, \underset{\sim}{\alpha}, \underset{\sim}{\beta}, \sigma^2)$ and $Z(\cdot) \sim AR(p, \underset{\sim}{\alpha}, \sigma^2)$ with associated covariance matrix sequences $\Gamma_{Y,t} = L_{Y,t} D_{Y,t} L_{Y,t}^T$ and $\Gamma_{Z,t} = L_{Z,t} D_{Z,t} L_{Z,t}^T$. Then

$$L_{Y,t} = L_{Z,t}L_{X,t} ,$$

$$D_{Y,t} = D_{X,t} ,$$

where $\underset{\sim}{X}_t = (X(1), \ldots, X(t))^T = L_{Z,t}^{-1}\underset{\sim}{Y}_t$ has jth element

$$X(j) = \begin{cases} Y(j) & , j = 1 \\ \sum_{k=0}^{j-1} \alpha_{j-1}(k)Y(j-k) & , j = 2, \ldots, p \\ \sum_{k=0}^{p} \alpha(k)Y(j-k) & , j > p \end{cases}$$

where the $\alpha_j(\cdot)$ are easily obtained by performing the Levinson (1974) recursion for $j = p-1, \ldots, 1$, with $\alpha_{p+1}(k) = \alpha(k)$:

$$\alpha_j(i) = \frac{\alpha_{j+1}(i) - \alpha_{j+1}(j+1)\alpha_{j+1}(j+1-i)}{1 - \alpha_{j+1}^2(j+1)} ,$$

$$i \le j < p.$$

Thus

$$\Gamma_{X,i,j} = \begin{cases} \sum_{m=\max(1,j-p)}^{j} \alpha_{j-1}(j-m) \\ \quad \times \sum_{\ell=\max(1,i-p)}^{i} \alpha_{i-1}(i-\ell)R_Y(\ell-m), & i,j \ge 1 \\ \\ R_X(|i-j|) = \sigma^2 \sum_{k=0}^{q-|i-j|} \beta(k)\beta(k+|i-j|), & \\ \qquad |i-j| \le q \quad i,j > p & \\ \\ 0 \quad \text{if } 1 \le j \le p,\ i > p, \text{ and } i - j > q & \\ \quad \text{or if } i,j > p \text{ and } |i-j| > q & \end{cases}$$

and

$$\lim_{K\to\infty} L_{X,K,K-j} = \beta(j), \quad j = 1, \ldots, q \qquad (3.7)$$

$$\lim_{K\to\infty} D_{X,K,K} = \sigma^2. \qquad (3.8)$$

Thus $e(1) = X(1)$, while

$$e(t) = X(t) - \sum_{k=1}^{\min(q,j-1)} L_{X,t,t-k}e(t-k), \quad t \ge 1$$

$$\sigma_t^2 = D_{X,t,t} ,$$

and in fact

$$Y(t+h|t) = X(t+h|t) - \sum_{j=1}^{p} \alpha(j)Y(t+h-j|t) ,$$

$$X(t+h|t) = \begin{cases} \sum_{k=h}^{q} L_{X,t+h,t+h-k}e(t+h-k), & \\ \qquad h = 1, \ldots, q & \\ 0 \qquad h > q & \end{cases}$$

$$\sigma_{t,h}^2 = \sum_{k=0}^{h-1} D_{X,t+h-k,t+h-k} \times \left\{ \sum_{\ell=t+h-k}^{t+h} L_{Z,t+h,\ell}L_{X,\ell,t+h-k} \right\}^2 .$$

Thus to find the one step ahead prediction errors and variances one need only compute the q nonzero elements of the successive rows of L_X until the convergence properties (3.7) and (3.8) take effect. Further, one can then also find the $Y(t+h|t)$ from these same quantities and the $e(\cdot)$. Note however that to find $\sigma_{t,h}^2$ one also needs to find the elements of $L_{Z,t}$. These can be found simply by noting that the upper (pxp) principal minor $L_{Z,p}$ can be found by inverting the lower triangular matrix $L_{Z,p}^{-1}$, while

$$L_{Z,j,k} = \gamma(j-k), \quad j \ge k \ge p ,$$

where $\gamma(0) = 1, \gamma(1), \gamma(2), \ldots$ are the coefficients in the MA(∞) representation of $Z(\cdot)$. Further, the elements in rows p+1, ... in the first (p-1) columns of L_Z are obtained by

$$L_{Z,p+j,k} = -\sum_{r=1}^{p} \alpha(r)L_{Z,p+j-r,k} , \quad 1 \le k < p, \quad j \ge 1,$$

and

$$\lim_{j\to\infty} L_{Z,p+j,k} = 0 , \qquad 1 \le k < p .$$

Thus the predictors and prediction variances $Y(t+h|t)$ and $\sigma_{t,h}^2$ can be obtained using either the Kalman filter algorithm or the Cholesky decomposition algorithm. Also the convergence properties described in the Cholesky algorithm can be incorporated into the Kalman filter algorithm.

4. DISCUSSION

Thus we have seen that algorithms developed recently for finding maximum likelihood estimators of ARMA process parameters have essentially consisted of applying established algorithms for finding the minimum mean square error linear one step ahead predictors and prediction variances. We have shown how the Cholesky algorithm can be used to find more than one step ahead predictors and variances as well. We note that both the Kalman filter and Cholesky decomposition algorithm can be extended easily to the multivariate ARMA case. We also note that we have not attempted to survey all the recent work in the ARMA estimation area, but rather have emphasized those algorithms that appear to be most widely used in the literature.

We consider next the relative speed and stability of the Kalman filter and Cholesky decomposition algorithms. Pearlman (1980) shows that

the number of operations needed to find the e(t) and σ_t^2 via the Kalman filter algorithm is approximately $T(2p + 3m + 3)$, with $m = \max(p,q+1)$, while for the Cholesky algorithm it is $T(p + \frac{1}{2}(q+1)(q+4))$. Thus if $q \geq 5$ the Kalman algorithm is faster.

Now the fact that the Kalman filter algorithm performs a number of "matrix squaring" operations (for example $H_t P_t H_t^T$ in (3.1) has caused many investigators to question its stability. However a number of authors have suggested methods for avoiding these operations (see Paige (1976) for example). The Cholesky decomposition is well known for its numerical stability (see Wilkinson (1967) for example). However more work is needed before a final conclusion can be made about which of the two methods is to be preferred.

Finally we point out that there are a variety of numerical methods for analyzing linear time series models in the work of control theorists particularly in a series of papers of Kailath (see the references in Aasnaes and Kailath (1973)) that need to be incorporated into the statistical literature.

5. REFERENCES

Aasnaes, H. B. and Kailath, T. (1973). An innovations approach to least squares estimation - part VII: some applications of vector autoregressive moving average models, IEEE trans on Automatic Control, AC-18, 601-607.

Akaike, H. (1973). Maximum likelihood identification of Gaussian autoregressive moving average models, Biometrika, 60, 255-265.

Akaike, H. (1974). Markovian representation of stochastic processes and its applications to the analysis of autoregressive moving average processes. Annals of the Institute of Statistical Mathematics, 26, 363-387.

Ansley, C. (1979). An algorithm for the exact likelihood of a mixed autoregressive-moving average process. Biometrika, 66, 59-65.

Bauer, F. L. (1955). Ein direktes iterationsverfahren zur Hurwitz-Zerlegung eines polynoms, Archiv Elekt. Ubertr., 9, 285-290.

Box, G. E. P. and Jenkins, G. M. (1970). Time Series Analysis forecasting and control. Holden-Day, San Francisco.

Dennis, J. E. and More, J. J. (1977). Quasi-Newton methods, motivation and theory. SIAM Rev., 19, 46-89.

Gardner, G., Harvey, A. C. and Phillips, G. D. (1980). An algorithm for exact maximum likelihood estimation of autoregressive-moving average models by means of Kalman filtering, JRSS C, 29, 311-322.

Hannan, E. J. (1969). The estimation of mixed moving average autoregressive systems. Biometrika, 56, 579-593.

Harvey, A. C. and Phillips, G. D. (1979). Maximum likelihood estimation of regression models with autoregressive-moving average disturbances. Biometrika, 66, 49-58.

Jones, R. H. (1980). Maximum likelihood fitting of ARMA models to time series with missing observations. Technometrics, 22, 389-395.

Levinson, N. (1974). The Wiener RMS error criterion in filter design and prediction. Journal of Mathematical Physics, 15, 261-278.

Newton, H. J. (1980). Efficient estimation of multivariate moving average autocovariances. Biometrika, 67, 227-231.

Newton, H. J. and Pagano, M. (1981). The finite memory prediction of covariance stationary time series. Texas A&M Statistics Technical Report N-21.

Pagano, M. and Parzen, E. (1973). Timesboard; A time series package. Proc. of Comp. Sci. and Stat.: 7th Ann. Symp. on the Interface, ed. by W. J. Kennedy, Statistical Laboratory, Iowa State Univ., Ames, Iowa.

Pagano, M. (1976). On the linear convergence of a covariance factorization algorithm. Journ. Assoc. Comp. Mach., 23, 310-316.

Paige. (1976). Numerical computations for some estimation problems in Engineering, Proc. of Comp. Sci. and Stat.: 9th Ann. Symp. on the Interface, Harvard University.

Pearlman, J. G. (1980). An algorithm for the exact likelihood of high-order autoregressive-moving average process. Biometrika, 67, 232-233.

Phadke, M. S. and Kedem, G. (1978). Computation of the exact likelihood function of multivariate moving average models. Biometrika, 65, 511-519.

Whittle, P. (1963). Prediction and Regulation by linear least squares methods. English Universities Press, London.

Wilkinson, J. H. (1967). The solution of ill-conditioned linear equations. In Mathematical Methods for Digital Computers II (A. Ralston and H. S. Wilf, eds.) 65-93.

Wilson, G. (1969). Factorization of the covariance generating function of a pure moving average process. SIAM J. Numer. Anal., 6, 1-7.

END

DATE
FILMED
10-81

DTIC